

CLAY: A Controllable Large-scale Generative Model for Creating High-quality 3D Assets

Year: 2024 | Citations: 293 | Authors: Longwen Zhang, Ziyu Wang, Qixuan Zhang, Qiwei Qiu, Anqi Pang

Abstract

In the realm of digital creativity, our potential to craft intricate 3D worlds from imagination is often hampered by the limitations of existing digital tools, which demand extensive expertise and efforts. To narrow this disparity, we introduce CLAY, a 3D geometry and material generator designed to effortlessly transform human imagination into intricate 3D digital structures. CLAY supports classic text or image inputs as well as 3D-aware controls from diverse primitives (multi-view images, voxels, bounding boxes, point clouds, implicit representations, etc). At its core is a large-scale generative model composed of a multi-resolution Variational Autoencoder (VAE) and a minimalistic latent Diffusion Transformer (DiT), to extract rich 3D priors directly from a diverse range of 3D geometries. Specifically, it adopts neural fields to represent continuous and complete surfaces and uses a geometry generative module with pure transformer blocks in latent space. We present a progressive training scheme to train CLAY on an ultra large 3D model dataset obtained through a carefully designed processing pipeline, resulting in a 3D native geometry generator with 1.5 billion parameters. For appearance generation, CLAY sets out to produce physically-based rendering (PBR) textures by employing a multi-view material diffusion model that can generate 2K resolution textures with diffuse, roughness, and metallic modalities. We demonstrate using CLAY for a range of controllable 3D asset creations, from sketchy conceptual designs to production ready assets with intricate details. Even first time users can easily use CLAY to bring their vivid 3D imaginations to life, unleashing unlimited creativity.